

## **BAB III**

### **DATA DAN PERANCANGAN SISTEM**

Bab ini menjelaskan mengenai dataset, perancangan alur kerja sistem, grafik data penelitian, dan perancangan pengujian. Adapun isi dari perancangan alur kerja sistem yaitu data pemakaian daya listrik, pemilihan data, pembagian data, perancangan pelatihan *Support Vector Regression* (SVR), evaluasi dan analisa. Dan isi dari perancangan pengujian yaitu perancangan pengujian dalam pengelolaan data.

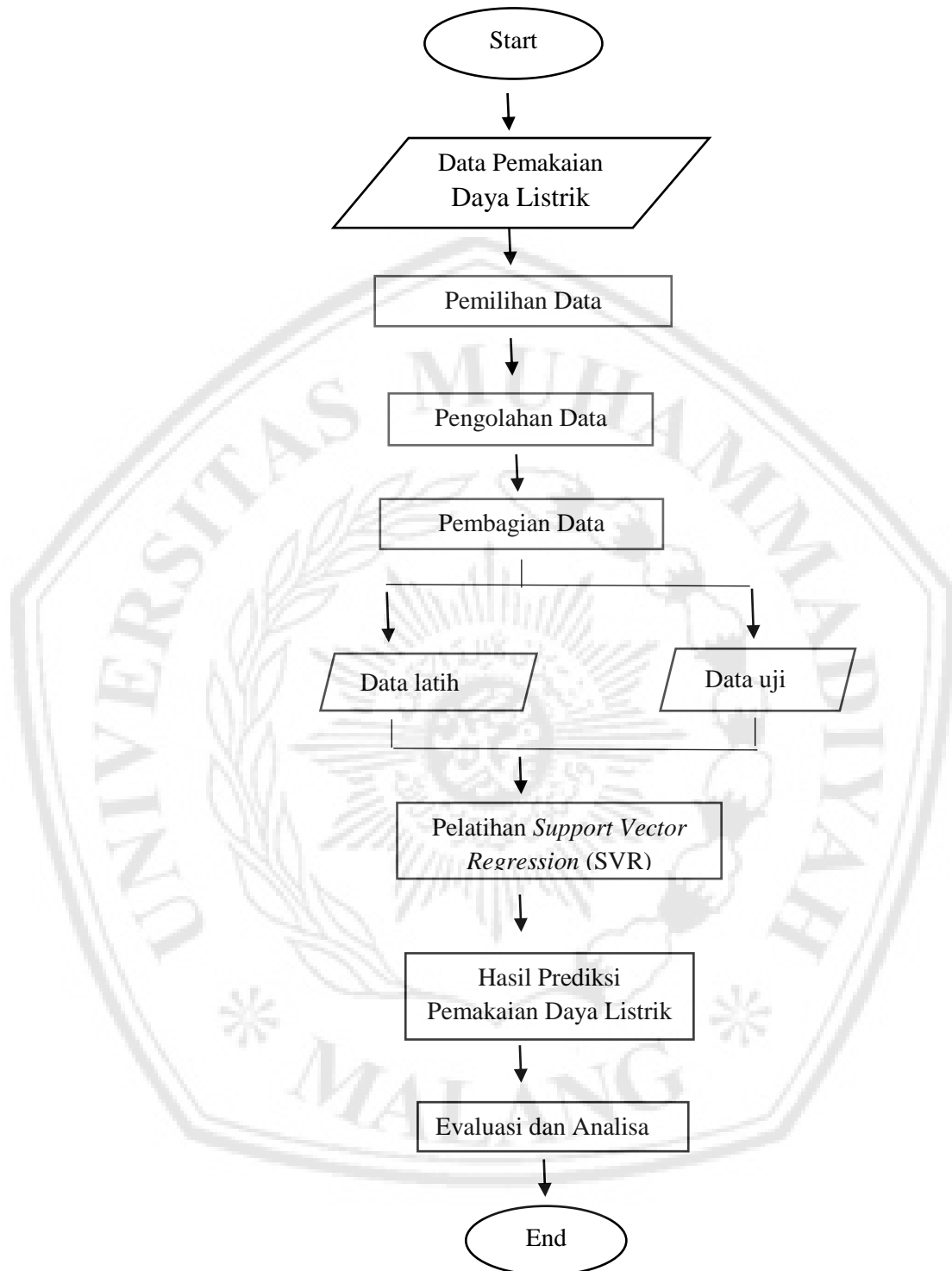
#### **3.1 Dataset**

Penelitian tugas akhir ini pengambilan data dilakukan di PT. PLN (Persero) Rayon Seririt, dimana data tersebut diperoleh dengan melakukan observasi di kantor PLN (Persero) Rayon Seririt, setelah mendapatkan ijin dari PH Manajer PT. PLN (Persero) Rayon Seririt, pada proses pengambilan data peneliti akan diarahkan ke bagian hubungan pelayanan pelanggan.

Obyek pengamatan adalah data pemakaian daya listrik untuk penentuan besarnya pemakaian daya listrik yang berjumlah 60 bulan periode 2012 dan sampai dengan 2016. Setelah semua data terkumpul dan diperoleh, maka ditentukan parameter dari data pemakaian daya listrik. Parameter data pemakaian daya listrik yang digunakan dalam algoritma *Support Vector Regression* (SVR) berupa jumlah pelanggan, pemakaian daya listrik. Setelah menentukan parameter, maka data dilakukan pengolahan data. Data akan dilakukan pengujian dengan algoritma *Support Vector Regression* (SVR) untuk mengetahui besarnya pemakaian daya listrik yang akan datang.

Data sebelum diimplementasi di java dilakukan preprocess menggunakan library weka untuk membuat model prediksi pemakaian daya listrik dengan algoritma *Support Vector Regression* (SVR). Pada penelitian ini menggunakan library Weka. Weka dalam penggunaannya menggunakan file ARFF (Attribute Relation File Format) adalah file yang digunakan Weka yang berisi 1 atau lebih instances dari attribut, maka data pemakaian daya listrik yang berformat arff.

### 3.2 Perancangan Alur Kerja Sistem



**Gambar 3.1.** Diagram Alur Kerja Sistem Prediksi Pemakaian Daya Listrik

### 3.2.1 Data Pemakaian Daya Listrik

Tahapan pertama dilakukan pengumpulan data pemakaian daya listrik yaitu dengan cara melakukan studi lapangan di PT. PLN (Persero) Rayon Seririt untuk mendapatkan data. Kemudian melakukan pengumpulan data yang diambil dari data observasi pemakaian daya listrik di PT. PLN (Persero) Rayon Seririt.

### 3.2.2 Pemilihan Data

Tahapan kedua yaitu setelah semua data terkumpul dan diperoleh, maka langkah selanjutnya dilakukan pemilihan data. Pemilihan data bertujuan untuk memilih parameter-parameter yang digunakan untuk pengolahan data pada tahap berikutnya. Parameter-parameter yang terpilih akan digunakan sebagai prediksi pemakaian daya listrik. Proses pemilihan parameter berdasarkan kriteria besarnya pemakaian daya listrik yang digunakan di PT. PLN (Persero) Rayon Seririt meliputi:

Kriteria 1 : Jumlah Pelanggan,

Kriteria 2: Pemakaian Daya.

Parameter algoritma yang digunakan dalam algoritma *Support Vector Regression* (SVR) berupa jumlah pelanggan, dan pemakaian daya. Setelah pemilihan parameter didapatkan hasil pemilihan parameter. Tabel 3.1 menunjukkan hasil pemilihan data pemakaian daya listrik bulan Januari 2014.

**Tabel 3.1** Hasil Pemilihan Data Pemakaian Daya Listrik Bulan Januari 2014

Jumlah Pelanggan	Pemakaian Daya
397	17.221
522	34.625
0	0
1.230	139.132
16.627	1.171.240
13.479	1.379.975
0	0
34.790	3.350.373
138	7.973
0	0
192	22.006

### 3.2.3 Pengolahan Data

Pengolahan data dilakukan dengan dua cara, yaitu *Preprocessing* data dan penjumlahan data. Tahap pertama adalah *preprocessing* data dilakukan untuk pembersihan data sehingga data siap untuk dilakukan regresi. Tahapan *Preprocessing* data pada penelitian ini adalah data *cleaning*. Proses data *cleaning* ini ditunjukkan pada Tabel 3.2.

**Tabel 3.2** Proses Data *Cleaning*

Pemakaian Daya		Pemakaian Daya
34.625		34.625
0		26.309
0		9.822
26.309	→	51.155
9.822		139.132
51.155		1.171.240
0		
139.132		
1.171.240		

Table diatas terdapat data yang kosong maka akan dilakukan data *cleaning*. Dalam data *cleaning* yang akan dilakukan yaitu menghapus data yang kosong atau data yang bernilai 0. Proses data *cleaning* yang dilakukan dengan menghapus semua parameter yang memiliki data yang kosong atau data yang bernilai 0.

Tahap kedua yaitu menjumlahkan data pada masing-masing parameter. Gambar 3.1 menunjukkan contoh penjumlahan parameter pemakaian daya bulan januari tahun 2014 sebagai berikut:  $17221 + 34625 + 26309 + 9822 + 51155 + 139132 + 1379975 + 559744 + 149687 + 53415 + 36312 + 7973 + 22006 + 48744 + 111662 + 287339 + 477724 + 1394 + 69121 + 70515 + 589 + 2062 + 1719 + 1978 + 52068 + 303992 + 362408 + 121461 + 168104 = 4568256$ .

Setelah proses penjumlahan selsai dilakukan maka akan didapatkan data pemakaian daya. Data pemakaian daya siap untuk diproses pada tahap selanjutnya. Jumlah data yang digunakan pada penelitian ini sebanyak 60 bulan. Tabel 3.3 menunjukkan hasil pengolahan data pemakaian daya listrik tahun 2014.

**Tabel 3.3** Hasil Pengolahan Data Pemakaian Daya Listrik Tahun 2014

Bulan	Jumlah Pelanggan	Pemakaian Kwh
Januari	37445	4568256
Februari	37610	4193155
Maret	37950	4588840
April	38427	4724674
Mei	38739	4951036
Juni	38953	4747148
Juli	39203	4807417
Agustus	39438	4811380
September	39749	4742244
Oktober	40119	5051849
November	40410	4896545
Desember	40675	5117621

#### 3.2.4 Pembagian Data

Dataset dibagi menjadi data latih dan data uji . Membagi keseluruhan data yang berjumlah 60 bulan menjadi data *training* dan data *testing*. Data *training* yang akan digunakan sebanyak 48 bulan dan data *testing* untuk prediksi sebanyak 12 bulan. Data latih atau data *train* dibuat sebagai data acuan untuk menentukan parameter dalam proses pengolahan terhadap data baru. Data latih berisi data pemakaian daya listrik di PT. PLN (Persero) Rayon Seririt. Data latih digunakan oleh algoritma *Support Vector Regression* (SVR) untuk membentuk sebuah model *Support Vector Regression* (SVR). Model ini merupakan representasi pengetahuan yang akan digunakan untuk prediksi pemakaian daya listrik. Contoh data latih pemakaian daya listrik dapat dilihat pada tabel 3.4.

**Tabel 3.4** Sample Data Latih Pemakaian Daya Listrik

Bulan	Jumlah Pelanggan	Pemakaian Daya
Januari	37445	4568256
Februari	37610	4193155
Maret	37950	4588840
April	38427	4724674
Mei	38739	4951036
Juni	38953	4747148
Juli	39203	4807417
Agustus	39438	4811380
September	39749	4742244
Oktober	40119	5051849
November	40410	4896545
Desember	40675	5117621

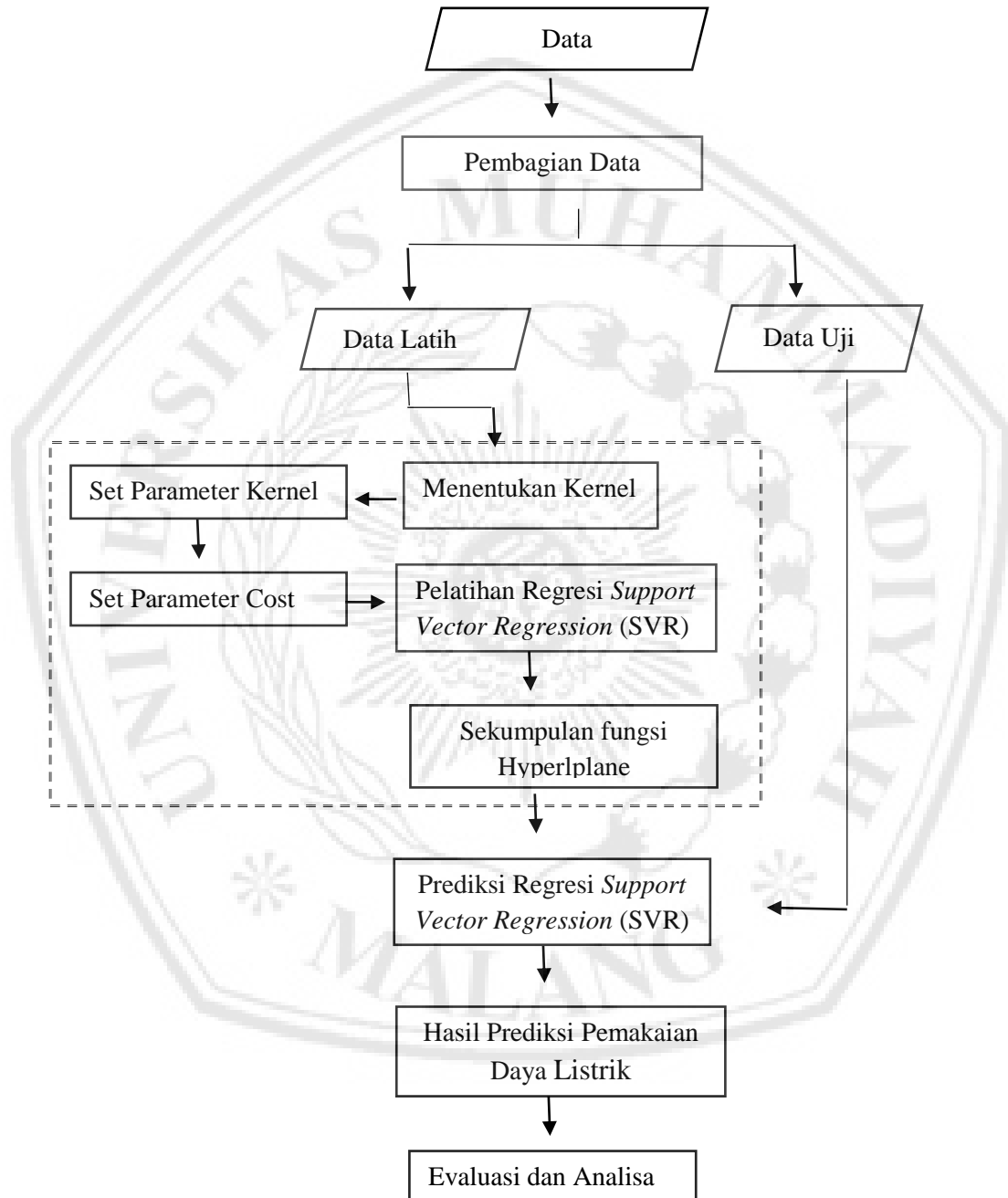
Data *test* atau data uji dibuat untuk melakukan uji dari data *train* (data latih). Data uji merupakan sebuah data yang digunakan untuk mengukur sejauh mana regresi *Support Vector Regression* (SVR) berhasil melakukan prediksi pemakaian daya listrik dengan benar. Data uji adalah bagian dari data latih namun proses penggunaannya berbeda. Data uji di gunakan untuk mengukur dan membandingkan hasil kinerja (evaluasi dan analisa) dari perhitungan algoritma *Support Vector Regression* (SVR). Contoh data latih pemakaian daya listrik dapat dilihat pada tabel 3.5.

**Tabel 3.5** Sample Data Uji Pemakaian Daya Listrik

Bulan	Jumlah Pelanggan	Pemakaian Daya
Mei	41668	4834086
Juni	42117	4873670
Juli	42302	5054738
Oktober	43050	5126676
November	43302	5125633
Desember	43621	5268093

### 3.2.5 Perancangan Pelatihan Regresi *Support Vector Regression* (SVR)

Proses pelatihan dimulai dengan menginputkan data hasil pengolahan data dahulu yang dapat dilihat pada proses pengolahan data dan pengujian kernel sesuai format software yang digunakan. Adapun langkah-langkah pelatihan prediksi *Support Vector Regression* (SVR) Gambar 3.2.



**Gambar 3.2.** Tahapan Perancangan Regresi *Support Vector Regression* (SVR)

Pelatihan data dilakukan dengan menggunakan 3 kernel, yaitu:

1. Kernel linear.
2. Kernel *polynomial*, membutuhkan parameter *degree* (d).
3. Kernel *Radial Basis Function* (RBF), membutuhkan parameter *gamma* ( $\gamma$ ).

Masing-masing kernel dicoba dengan nilai parameter fungsi kernel, hal ini dilakukan pada fungsi kernel *polynomial* dan kernel *Radial Basis Function* (RBF) terkecuali kernel linear. Sehingga didapat beberapa model *Support Vector Regression* (SVR) yang masing-masing akan diuji dan akan dihasilkan nilai *output* dari masing-masing model regresi yang dapat menentukan prediksi pemakaian daya listrik yang akan datang.

### **3.2.6 Hasil Prediksi Pemakaian Daya Listrik**

Tahapan ini yaitu setelah dilakukan proses pelatihan *Support Vector Regression*, maka akan mendapatkan model *Support Vector Regression* (SVR) yang masing-masing akan diuji dan akan dihasilkan nilai *output* dari masing-masing model regresi yang dapat menentukan prediksi pemakaian daya listrik yang akan datang. Sehingga didapatkan hasil prediksi pemakaian daya listrik yang akan datang.

### **3.2.7 Evaluasi dan Analisa**

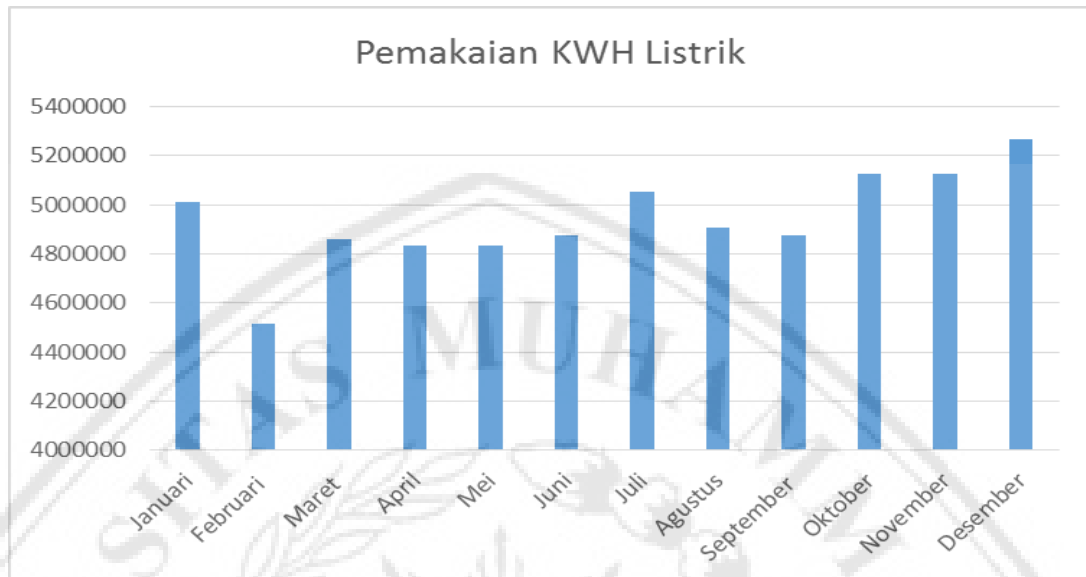
Tahap terakhir adalah tahap analisa, dimana pada tahap ini dilakukan analisa terhadap pengukuran akurasi hasil prediksi yang diperoleh dengan model *Support Vector Regression* (SVR) terhadap data uji untuk mempresentasikan hasil informasi yang ditentukan.

## **3.3 Grafik Data Penelitian**

Grafik adalah alat penyajian statistik yang tertuang dalam bentuk lukisan, baik lukisan garis, lukisan gambar, maupun lambang [7]. Grafik juga dapat diartikan sebagai rangka yang dapat dibentuk menjadi objek visualisasi tabel. Tabel yang ini terdiri dari angka yang akan disajikan atau dapat pula diberikan tampilan dengan bentuk gambar, garis, lingkaran, dan lain-lain. Grafik menjadi gambaran tentang pasang surut akan suatu keadaan atau data yang akan di gambarkan dengan gambar atau baris. Grafik dibagi menjadi 3 bagian yaitu lingkaran, batang,



dan garis. Grafik di buat dengan tujuan yaitu untuk dapat menunjukkan perbandingan informasi yang berkualitas dengan cepat dan sederhana [7]. Contoh grafik data pemakaian daya listrik bulan januari tahun 2016.



**Gambar 3.3** Grafik Data Pemakaian Daya Listrik Bulan Januari 2016

Gambar 3.1 menunjukkan grafik data pemakaian daya listrik bulan januari 2016, dimana Sumbu data (sumbu x) menyatakan bulan dan sumbu tegak vertikal menyatakan pemakaian daya listrik. Tampak grafik mengalami fluktuasi dari bulan januari sampai dengan desember. Berdasarkan tabel diatas, jumlah pemakaian daya listrik mencapai angka tertinggi terdapat pada bulan desember. Sedangkan pemakiaandaya listrik mencapai angka terendah pada bulan februari. Parameter pemakaian daya listrik digunakan untuk mengetahui besarnya penggunaan daya listrik tiap bulan.

### 3.4 Perancangan Pengujian

#### 3.4.1 Perancangan Pengujian Regresi *Support Vector Regression* (SVR)

Pengujian dalam penelitian data train dan data test sama dilakukan proses pengolahan data dahulu yang dapat dilihat pada Tabel 3.2 dan Tabel 3.3. Pembagian parameter yang digunakan pada tahap pengujian data train dan data test diambil keseluruhan parameter yaitu jumlah pelanggan, pemakaian daya listrik. Data akan diuji dengan menggunakan pengujian validasi.

Pengujian validasi dilakukan untuk mengetahui keberhasilan sistem regresi. Sebelum dilakukan pengujian validasi terlebih dahulu akan dilakukan pengujian pelatihan pemakaian daya listrik. Data latih akan dilakukan pengujian pelatihan daya listrik menggunakan algoritma *Support Vector Regression* (SVR). Setelah dilakukan proses pelatihan *Support Vector Regression*, maka akan mendapatkan model *Support Vector Regression* (SVR) yang masing-masing akan diuji dengan data uji dan akan dihasilkan nilai *output* dari masing-masing model regresi yang dapat menentukan prediksi pemakaian daya listrik yang akan datang.

Data uji digunakan untuk mengevaluasi kinerja model *Support Vector Regression* (SVR) yang telah terbentuk sebelumnya dengan cara data uji dimasukkan pada model *Support Vector Regression* (SVR) yang telah terbentuk. Untuk mengevaluasi data uji dengan model regresi menggunakan proses pengujian validasi pada data uji. Setelah melalui proses prediksi menggunakan algoritma *Support Vector Regression* (SVR) dan output yang dihasilkan berupa angka numerik. Kemudian dengan hasil numerik tersebut dilakukan pengujian validasi nya dapat menggunakan sejumlah indikator mengukur kesalahan error. Proses pelatihan ini untuk pengujian validasi memilih menggunakan kesalahan error *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), *Koefisien Korelasi* (R). Proses aelanjutnya adalah data prediksi pemakaian daya listrik yang dihasilkan oleh model *Support Vector Regression* (SVR) akan dibandingkan hasil akurasi ukuran error nilai fungsi kernel linier, kernel polynomial, kernel *Radial Basis Function* (RBF) dan ditentukan mana hasil akurasi ukuran error nilai fungsi kernel dari ketiga kernel yang paling bagus. Berikut merupakan perhitungan nilai akurasi *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), *Koefisien Korelasi* (R):

1. *Mean Absolute Error* (MAE)

$$\begin{aligned}
 & \sum_{i=1}^{12} \left| (4749748,3 - 5009460,0) + (4784957,1 - 4517661,0) + \right. \\
 & \quad \left. (4810243,5 - 4861098,0) + (4833769,4 - 4833019,0) + \right. \\
 & \quad \left. (4863376,8 - 4834086,0) + (4935234,9 - 4873670,0) + \right. \\
 & \quad \left. (4964842,3 - 5054738,0) + (5008853,4 - 4905275,0) + \right. \\
 & \quad \left. (5049983,7 - 4874014,0) + (5084552,4 - 5126676,0) + \right. \\
 & \quad \left. (5124882,5 - 5125633,0) + (5175935,3 - 5268093,0) \right| \\
 &= \frac{\quad}{12} \quad (12) \\
 &= \frac{1173944}{12} \\
 &= 97828,66
 \end{aligned}$$

2. *Root Mean Squared Error (RMSE)*

$$\begin{aligned} &= \sqrt{\frac{1}{12} (206175363078,79)} \\ &= \sqrt{17181280256,566} \\ &= 131077,3 \end{aligned} \tag{13}$$

3. Koefisien Korelasi

$$\begin{aligned} &= \frac{(2936E+14)-(293289E+14) (294117E+14)}{\sqrt{((293289E+14)-(59283423))^2 ((294117E+14)-(59386379,6))^2}} \\ &= 66.4 \% \end{aligned} \tag{14}$$

